

TAREA 1: RECUPERACIÓN DE TWEETS BASADO EN EL CONTENIDO

- Implemente el índice invertido usando el modelo de recuperación por ranking para consultas de texto libre.
 - o Use la similitud de coseno sobre los pesos TF-IDF.
 - o La consulta es solo una o más palabras en lenguaje natural.
- Para probar el desempeño de su implementación. Se proveerá una colección de aproximadamente 20mil tweets de Twitter.
 - o El diccionario de términos se construye usando el contenido del atributo "text". El docID vendría a ser el Id del tweet.
 - o Se debe filtrar los stopwords encontrados.
- Proponga tres consultas y muestre el **top-10 de los tweets** que se aproximan a dicha consulta.
- Analice el performance de su implementación y proponga una la solución algorítmica para el uso de memoria secundaria ante grandes colecciones de datos.

Ejemplo de tweets:

```
[
  {
    "id": 1046263368691023873,
    "date": "Sun Sep 30 05:00:00 +0000 2018",
    "text": "#VotaBien @EstherCapunay Necesito esa via expresa sur pero yaaaaaa !!!",
    "user_id": 1042868076327231488,
    "user_name": "@AlcantaraYasuri",
    "location": {},
    "retweeted": false
  },
  {
    "id": 1046263372675788800,
    "date": "Sun Sep 30 05:00:01 +0000 2018",
    "text": "RT @PeruanoComunica: Jorge Muñoz era la voz, hasta que vi esta foto. Vitocho el ultra KeikoAlanista que recibirá ordenes de la #SeñoraK y #...",
    "user_id": 861714162979934209,
    "user_name": "@Emperilluminati",
    "location": {},
    "retweeted": true,
    "RT_text": "Jorge Muñoz era la voz, hasta que vi esta foto. Vitocho el ultra KeikoAlanista que recibirá ordenes de la #SeñoraK y #AG. Ni que hablar de los otros dos. 🇵🇪 NO PODEMOS DAR NI UN MILÍMETRO DE PASO A KEIKO CON MIRAS AL 2021 https://t.co/DGT8QBw0Z1",
    "RT_user_id": 2977539507,
    "RT_user_name": "@PeruanoComunica"
  },
]
```